

Contingency Analysis.R

Administrator

Tue Apr 05 15:00:55 2016

```
# Analysis of simple contingency tables
# 5 April 2016
# NJG
```

```
# Create a simple data table
```

```
vec1 <- c(50,66,22)
vec2 <- c(120,22,30)
data <- rbind(vec1,vec2)
rownames(data) <- c("Cold","Warm")
colnames(data) <- c("Aphaenogaster", "Camponotus", "Crematogaster")
print(data)
```

```
##      Aphaenogaster Camponotus Crematogaster
## Cold              50          66           22
## Warm             120          22           30
```

```
# Create a simple data table
```

```
vec1 <- c(50,66,22)
vec2 <- c(120,22,30)
data <- rbind(vec1,vec2)
rownames(data) <- c("Cold","Warm")
colnames(data) <- c("Aphaenogaster", "Camponotus", "Crematogaster")
print(data)
```

```
##      Aphaenogaster Camponotus Crematogaster
## Cold              50          66           22
## Warm             120          22           30
```

```
# Sometimes the data table will have 1 row for each observation, with different columns for the factors
vec1 <- rep(c("Aphaenogaster","Camponotus","Crematogaster"),times=c(10,5,20))
print(vec1)
```

```
## [1] "Aphaenogaster" "Aphaenogaster" "Aphaenogaster" "Aphaenogaster"
## [5] "Aphaenogaster" "Aphaenogaster" "Aphaenogaster" "Aphaenogaster"
## [9] "Aphaenogaster" "Aphaenogaster" "Camponotus"      "Camponotus"
## [13] "Camponotus"    "Camponotus"    "Camponotus"    "Crematogaster"
## [17] "Crematogaster" "Crematogaster" "Crematogaster" "Crematogaster"
## [21] "Crematogaster" "Crematogaster" "Crematogaster" "Crematogaster"
## [25] "Crematogaster" "Crematogaster" "Crematogaster" "Crematogaster"
## [29] "Crematogaster" "Crematogaster" "Crematogaster" "Crematogaster"
## [33] "Crematogaster" "Crematogaster" "Crematogaster"
```

```
vec2 <- rep(c("Cold","Warm"),times=15)
data <- as.data.frame(cbind(vec1,vec2))
```

```
## Warning in cbind(vec1, vec2): number of rows of result is not a multiple of
## vector length (arg 2)
```

```
print(data)
```

```
##           vec1 vec2
## 1 Aphaenogaster Cold
## 2 Aphaenogaster Warm
## 3 Aphaenogaster Cold
## 4 Aphaenogaster Warm
## 5 Aphaenogaster Cold
## 6 Aphaenogaster Warm
## 7 Aphaenogaster Cold
## 8 Aphaenogaster Warm
## 9 Aphaenogaster Cold
## 10 Aphaenogaster Warm
## 11 Camponotus Cold
## 12 Camponotus Warm
## 13 Camponotus Cold
## 14 Camponotus Warm
## 15 Camponotus Cold
## 16 Crematogaster Warm
## 17 Crematogaster Cold
## 18 Crematogaster Warm
## 19 Crematogaster Cold
## 20 Crematogaster Warm
## 21 Crematogaster Cold
## 22 Crematogaster Warm
## 23 Crematogaster Cold
## 24 Crematogaster Warm
## 25 Crematogaster Cold
## 26 Crematogaster Warm
## 27 Crematogaster Cold
## 28 Crematogaster Warm
## 29 Crematogaster Cold
## 30 Crematogaster Warm
## 31 Crematogaster Cold
## 32 Crematogaster Warm
## 33 Crematogaster Cold
## 34 Crematogaster Warm
## 35 Crematogaster Cold
```

```
print(chisq.test(x=data$vec1,y=data$vec2))
```

```
## Warning in chisq.test(x = data$vec1, y = data$vec2): Chi-squared
## approximation may be incorrect
```

```
##
## Pearson's Chi-squared test
##
## data: data$vec1 and data$vec2
## X-squared = 0.17157, df = 2, p-value = 0.9178
```

```
# Estimate the standard deviation and error of sample proportion data
```

```
# standard deviation of proportion = sqrt(p*(1-p)/n)
```

```
Success <- 22
```

```
Failure <- 1
```

```
n <- Success + Failure
```

```
prob <- Success/n
```

```
sd_prob <- sqrt(prob*(1-prob)/n)
```

```
# but what does a confidence interval look like?
```

```
# if we use the normal, we get
```

```
qnorm(p=c(0.025,0.975,0.50),mean=prob,sd=sd_prob)
```

```
## [1] 0.8731790 1.0398645 0.9565217
```

```
# better to use a binomial distribution
```

```
qbinom(p=c(0.025,0.975,0.50),size=33,prob=prob)
```

```
## [1] 29 33 32
```

```
# or a beta for the distribution of probabilities
```

```
qbeta(p=c(0.025,0.975,0.50),shape1=22,shape2=1)
```

```
## [1] 0.8456275 0.9988499 0.9689845
```